

SYSTEM IMPROVEMENT AND REPORTING DIVISION

BEYOND MIRS DATA TECHNICAL REPORT

FEBRUARY 2005

ATA LIBRARY
11010 - 142 Street NW
Edmonton, AB
T5N 2R1

Alberta
EDUCATION

~~AE
BEY~~



ALBERTA EDUCATION CATALOGUING IN PUBLICATION DATA

Alberta. Alberta Education. System Improvement and Reporting Division.
Beyond MIRS data technical report.

ISBN 0-7785-3783-8

1. Educational evaluation. 2. Program evaluation. 3. Evaluation studies.
4. Educational tests and measurements. I. Title.

LB2822.75.A333 2005

379.1

For further information, contact

System Improvement and Reporting Division
9th Floor, Commerce Place
10155 102 Street
Edmonton, Alberta T5J 4L5

Telephone (780) 422-8671
Toll free in Alberta by dialing 310-0000
Fax: (780) 422-8345
Email: SIG@Edc.Gov.ab.ca

This document is intended primarily for:

System and School Administrators
Alberta Education Executive Team and Managers

And may be of interest to:

Teachers
Parents
Education Stakeholders
Community Members

Copyright © 2005, the Crown in Right of Alberta, as represented by the Minister of Education.

Permission is given by the copyright owner to reproduce this document for educational purposes and on a non-profit basis.

Table of Contents/Outline

Table of Contents/Outline	i
Introduction.....	1
Description of data	1
Limitations of the Data	3
Descriptive Statistics of Data.....	4
Discussion.....	11
GLA and PAT by Age within Grade Cohorts.....	12
Discussion.....	15
GLA by PAT- <i>Comparisons using achievement levels</i>	16
Gamma.....	16
Analysis of Students Below Grade Level	17
Discussion.....	19
GLA and Gender.....	20
Discussion.....	21
Data Availability for Sub-Groups of Coded Students	22
Conclusions.....	23
Bibliography	24

Introduction

Beyond MIRS: New Directions for Program Evaluation, detailed proposed revisions to the Ministry's approach to program evaluation and a foundation for a pilot project launched in September 2003 to assess the viability and utility of using classroom based student assessment data in service of program evaluations. The Technical Report is one in a series of reports¹, which individually examine aspects of the Beyond MIRS, but when read in conjunction are meant to provide a complete understanding of the project as a whole. This report attempts to ascertain the utility, as prescribed in the Beyond MIRS pilot project backgrounder, of the classroom based grade level of achievement (GLA) data collected. The two main areas for inquiry are:

1. Can Grade Level of Achievement data driven by formative classroom assessment methods be a reasonable approach, with acceptable concurrent and predictive validity, for generating data for judging program impacts?
2. Can Grade Level of Achievement data add value to student reporting mechanisms already in place in schools and support related processes of critical reflection at the classroom and school levels, and does GLA aid in designing revisions to the jurisdiction and provincial student information systems?

Description of data

202 schools from 4 jurisdictions submitted grade level of achievement data for 51,816 students². The fields collected included student name (surname and given name), Alberta student number, and enrolled grade. Enrolled Grade was defined as the grade to which the student was assigned. Typically there is a strong relationship between a student's age, peer group and enrolled grade.

GLA was collected for all students on graded curriculum, including those with special needs, in the following fields where applicable:

- GLA in English Language Arts
- If applicable (FL1 or FL2 students) - GLA in French Language Arts
- GLA in Mathematics

Grade Level of Achievement was defined as the grade level expressed as a whole number in relationship to the learning outcomes defined in the Program of Studies that teachers judged the student to have achieved at the end of the 2003/04 school year. The following examples were provided as guidelines:

¹ The other reports in the series include the project backgrounder titled *Beyond MIRS- New directions for program evaluation*, the Beyond MIRS Pilot Project Assessment, and the project description paper titled *Accountability for Learning- The Beyond MIRS Project*.

² The majority of the data (98.2%) was submitted by the Edmonton Public School District. This number represents the total number of student records and includes those where GLA data was missing.

- *If the student is in Grade 1 and, if you judge that as of the end of the current school year he/she has met the learning outcomes in the Program of Studies for grade 1 Language Arts you would indicate "achieved grade 1"; and if you judge the student has not met the learning objectives in Math you would indicate "not yet 1."*
- *If the student is in Grade 1 and is performing above grade level, record the grade level at which you judge the student is performing, e.g. "achieved grade 3."*
- *For students in FL1 programs (students entitled under Section 23 of the Charter of Rights and Freedom to enroll in French first language schools) and in FL2 (French Immersion programs) report both a grade level of achievement for the student's French Language Arts program, and for their English Language Arts program at the end of the year that English Language Arts instruction is initiated (may range from grade 1 to 3).*
- *For students in FL1 and in FL2 programs report a single grade level of achievement for Mathematics independent of the language of Math instruction.*

Some school boards apply a standard test or battery of tests to determine the grade level of achievement. If that was true for the school submitting the data, teachers were asked to consider that assessment in relationship to the full range of assessment information available to them, including classroom assessment marks, in making a professional judgment of the student's grade level of achievement.

For students not on a graded curriculum (i.e. not based on the Programs of Study), teachers were asked to check one of the following descriptions that best described the goals in the student's Individualized Program Plan (IPP) that had been met. If goals were met, teachers were asked to respond, "YES". If the goals were not met teachers were instructed to respond "NO", and if not applicable they were instructed to respond "N/A".

- *Student has met IPP goals and objectives that address communication skills.*
- *Student has met IPP goals and objectives that address functional skills.*

Not on Graded Curriculum was meant to indicate that the student's program was restricted to learning outcomes that were significantly different from the provincial curriculum defined in the Program of Studies and were specifically selected to meet the student's special needs as defined in the *Standards for Special Education* (2002).

Communication skills referred to the development of expressive and or receptive communication. This could be verbal communication and/or alternative modes of communication.

Functional skills referred to skills that would assist the student in developing independence in the home, school and community.

The following illustrative examples were provided to help increase the reliability of the data received:

- Student A is enrolled in grade 4. Her Language Arts program is based on the grade 4 learning outcomes defined in the English Language Arts k-9 Program of Studies. The full range of assessment results for Student A demonstrates she has achieved the outcomes for grade 4 so the data is entered, “achieved grade 4.”
- Student B is enrolled in grade 8. He has been coded as having a mild learning disability. His Math program is based on the grade 6 learning outcomes defined in the Math k-9 Program of Studies. The full range of assessment results for Student B demonstrates he has achieved the outcomes for grade 6 so the data is entered, “achieved grade 6.”
- Student C is enrolled in grade 2. He has been coded as having a severe learning disability. His Language Arts program is based on developing language arts readiness skills and on some of the grade 1 learning outcomes defined in the English Language Arts k-9 Program of Studies. The full range of assessment results for Student C demonstrates he has not achieved all of the learning outcomes for grade 1 so the data is entered, “not yet 1.”
- Student D is enrolled in grade 3. She has been coded as having multiple severe disabilities and works with a full time aide. Her program is based completely on learning objectives that are below the grade 1 learning outcomes defined in the Math or English Language Arts k-9 Program of Studies. Her Individualized Program Plan defines communication and functional skill outcomes designed to develop independent living skills. All of the IPP outcomes for the current school year have been achieved so the data is entered in Part C, “Yes” for both communication and functional skills.

The Alberta student number was used by Alberta Education staff to append data fields such as Provincial Achievement Test (PAT) results (both raw scores and achievement levels), student age, number of school registrations, any additional student learning codes associated with the student, and starting date. Individual student identifiers were replaced with a discreet Beyond MIRS ID, leaving no personal identifiers in the dataset. (See Appendix 1 in Full Technical Report for a complete list of variables and descriptors.)

Limitations of the Data

When analyzing the data, the following limitations were noted.

- Nearly 98% of the data submitted was from one jurisdiction, which has been collecting GLA data for approximately 8 years.
- Of the total 51,816 records, 1,456 (approximately 2.8%) had no GLA data submitted for English Language Arts, and 1,358 (approximately 2.6%) had no GLA data submitted for Math.
- Of the 934 records submitted by other jurisdictions, 69 of the records submitted had no English Language Arts GLA data. However, 57 of these had IPP data submitted, meaning there was only 1.5% of the valid population with no English Language Arts GLA data. 62.3% of the same population had no data submitted for Math GLA.
- IPP data were submitted for only 57 students, meaning there were only 57 students not on a graded curriculum.

Descriptive Statistics of Data

The data are roughly evenly distributed by enrolled grade with 10-11% of the overall students from grade 1 to grade 9 in each grade cohort. If the students were distributed exactly evenly in each grade, we would expect 5671 students per grade, or approximately 11%. The table below shows the distribution by enrolled grade.

<i>Enrolled Grade Distribution</i>		
Enrolled Grade	Frequency	Valid Percent
1	5228	10.2%
2	5385	10.6%
3	5559	10.9%
4	5661	11.1%
5	5711	11.2%
6	5831	11.4%
7	6272	12.3%
8	6099	11.9%
9	5292	10.4%
Sub-Total	51038	100.0%
10	778	
Total	51816	

An irregularity was apparent in that there were 778 students in the database with an enrolled grade of 10. As the data was collected only for students in grades 1 to 9, these 778 were treated as anomalies and not used in any analyses by enrolled grade. They were however used as valid cases in analyses that were not grade specific.

88.9% of the students were Non-Coded, meaning their group code designation was "100". Approximately 2.1% (1,080) of the students had severe codes (codes 40 through 49), 7.8% (4066) had mild/moderate codes, and 1.2% were coded as gifted/talented (609).

<i>Recoded Expanded Code Variable into Groups- Population Parameters</i>		
	Frequency	Percent
Non-Coded	46061	88.9%
Severe Disabilities (Code 40 thru 49)	1080	2.1%
Mild/Moderate Disabilities	4066	7.8%
Gifted and Talented	609	1.2%
Total	51816	100%

Correlations between GLA and Enrolled Grade by Sub-Groups of the Population-

Correlations between the students' GLAs and enrolled grades were calculated using Spearman's rho to determine the "goodness of fit" between GLA and enrolled grade. The correlation between the two variables reflects the degree to which the variables are related, or the degree to which they "move" together. A high positive correlation coefficient results when an increase in one variable is mirrored by the same increase in another. Spearman's is used specifically to measure ordinal level data in that it first converts the data to rank orders before correlating. As was expected, GLA was highly correlated to enrolled grade, meaning the enrolled grade of a student matches their GLA. This relationship was strongest for the sub-population that had no group codes attached to their records, or non-coded students, while students with severe disabilities had the lowest correlation between GLA and enrolled grade. This relationship was true when testing both Math and English Language Arts GLA against enrolled grade.

English GLA

<i>Recoded Expanded codes into groups</i>	<i>Correlation Coefficient</i>
Non-Coded	.992(**)
Severe Disabilities (Code 40 thru 49)	.788(**)
Mild/Moderate Disabilities	.852(**)
Gifted and Talented	.999(**)

Math GLA

<i>Recoded Expanded codes into groups</i>	<i>Correlation Coefficient</i>
Non-Coded	.995(**)
Severe Disabilities (Code 40 thru 49)	.857(**)
Mild/Moderate Disabilities	.868(**)
Gifted and Talented	.974(**)

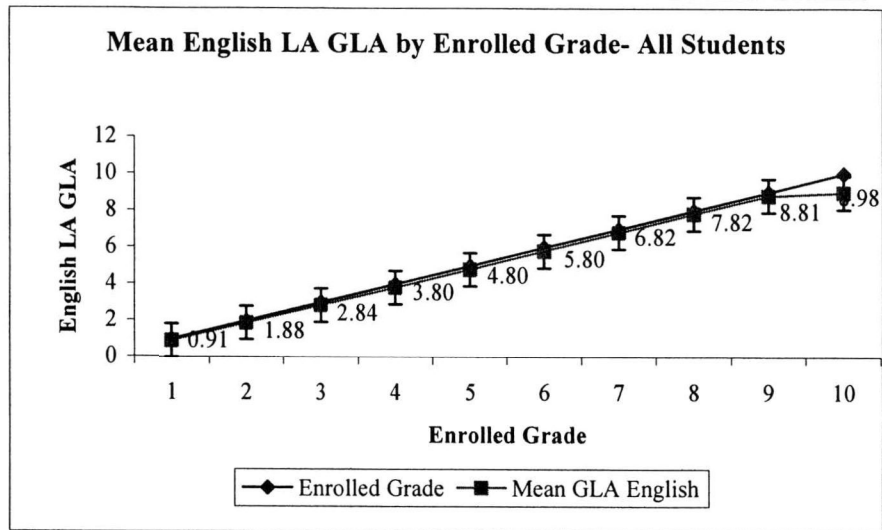
** Correlation is significant at the 0.01 level (2-tailed).

The following graphs (next page) show GLA by enrolled grade for all students as well as sub-populations of students, in English Language Arts and Math.

The mean GLA is plotted against the enrolled grade to show the degree to which students' GLA reflect their enrolled grade. Additionally, a trendline was plotted for each graph using the formula $y = bx + a$, where y is the dependent variable, b is the slope, x is the independent variable and a is the y-intercept, or the value at which the line would cross the y-axis. B indicates the amount of increase in the dependent variable when the independent variable is increased by 1.

In the "Mean English LA GLA by Enrolled Grade- All Students" graph, the slope (b) is .9546, meaning we can expect GLA to increase by roughly .95 when the enrolled grade is increased by 1. In other words, this is a nearly perfect positive correlation. For complete frequency tables of GLA compared to enrolled grade see Appendix 2 of the Full Technical Report.

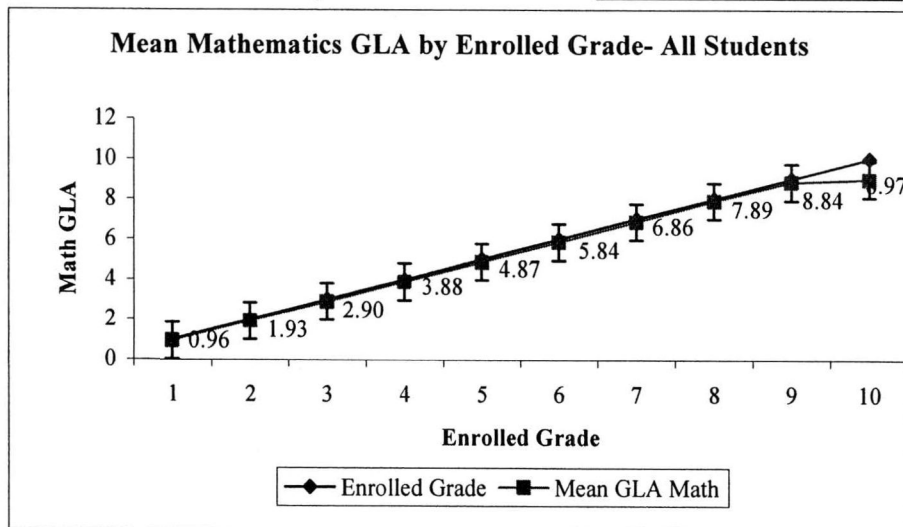
All Students- Entire Beyond MIRS Database ENGLISH LA GLA



Above mean GLA line is $y = 0.9546x + 0.0435$

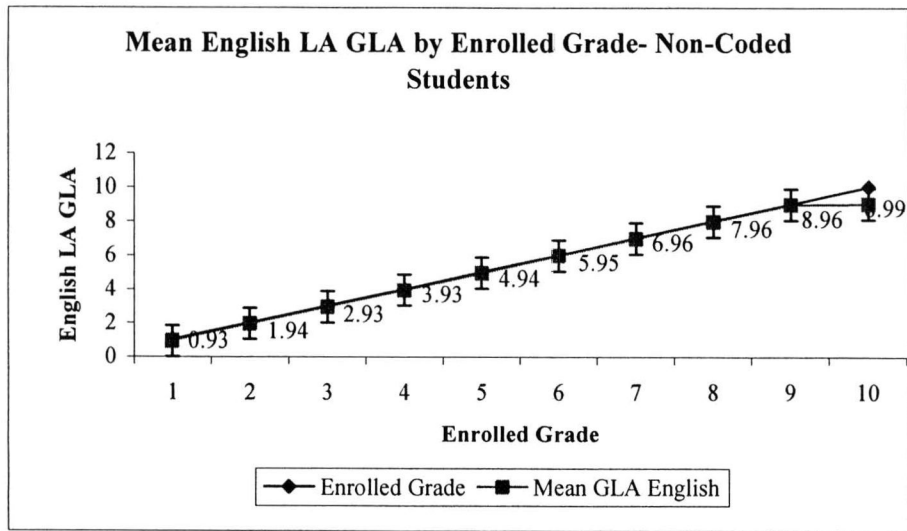
Formula for line is $y = bx + a$

All Students- Entire Beyond MIRS Database MATHEMATICS GLA



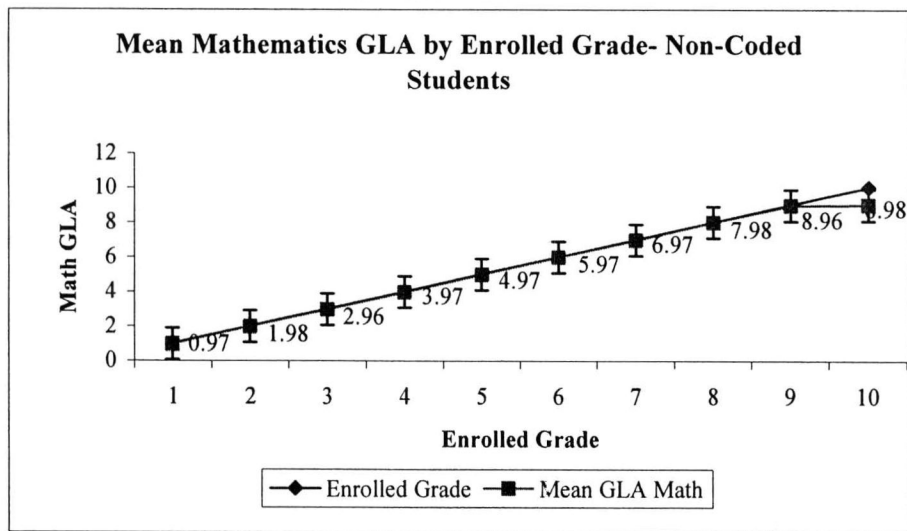
Above line is $y = .9412 + 0.1159$

Non-Coded Students



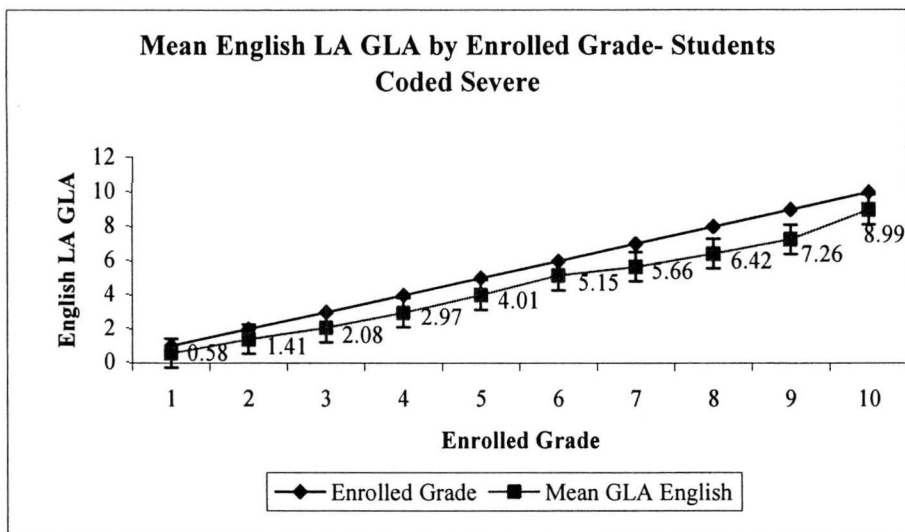
Above mean GLA line is $y = .9512x + 0.1175$

Non-Coded Students



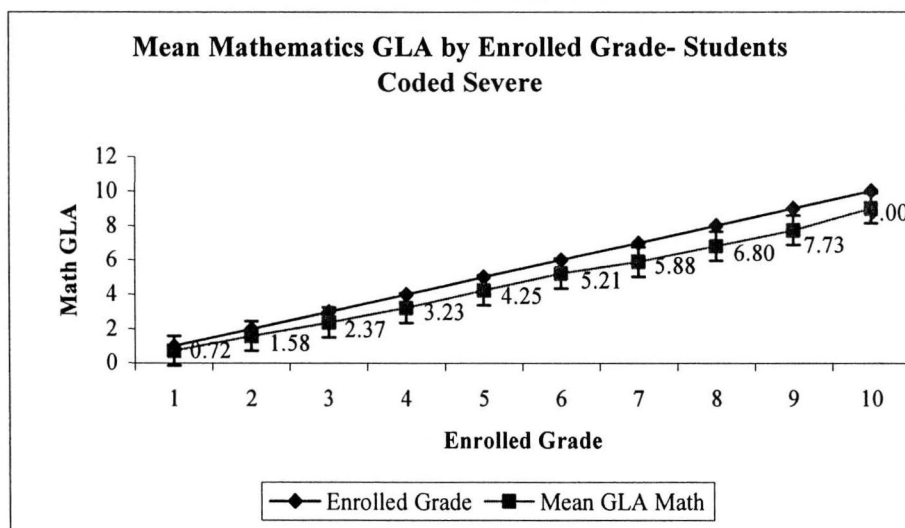
Above Mean GLA Math line is $y = .9457x + 0.171$

Students with Severe Codes



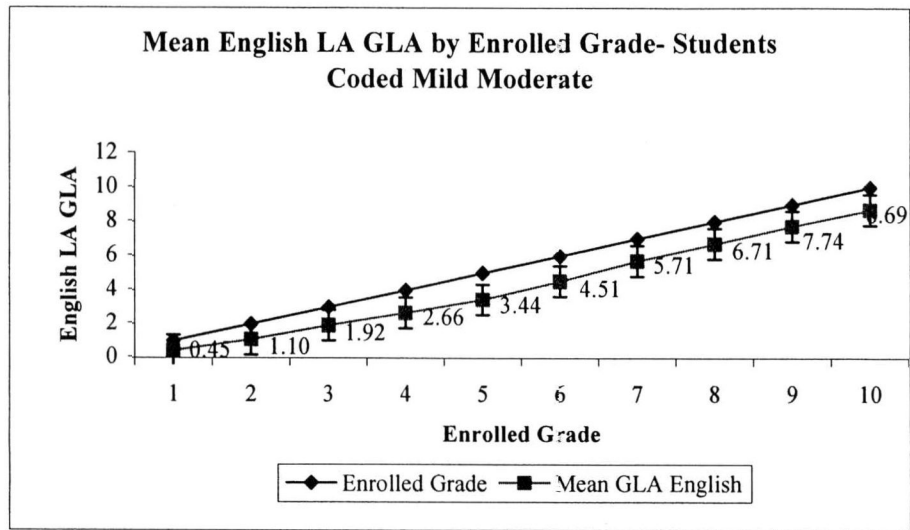
Above mean GLA line is $y = .8946x - 0.465$

Students with Severe Codes



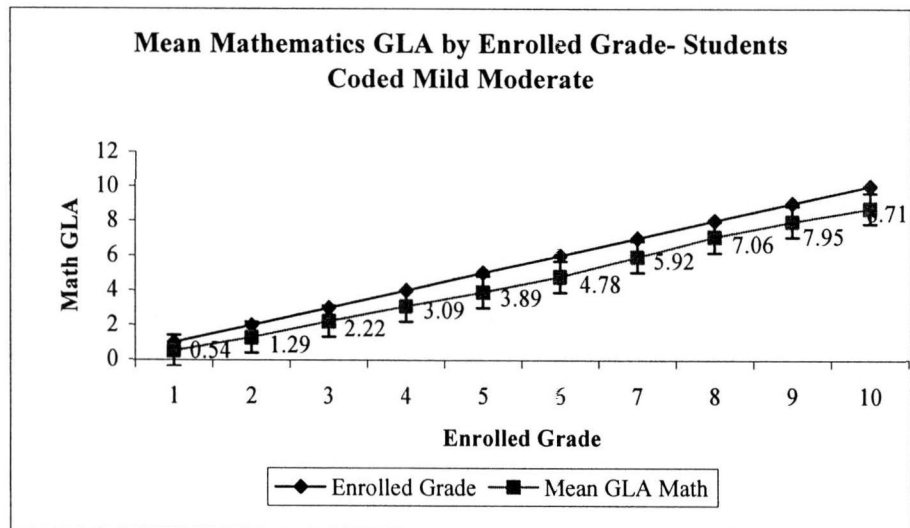
Above line for mean Math GLA is $y = .9008x - 0.2764$

Students with Mild Moderate Codes



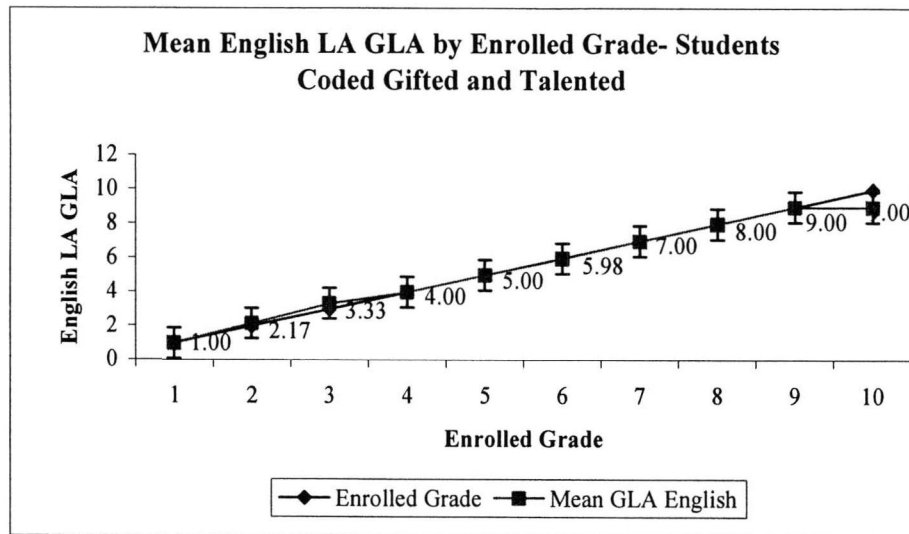
Above mean GLA line is $y = .9385x - 0.8692$

Students with Mild/Moderate Codes



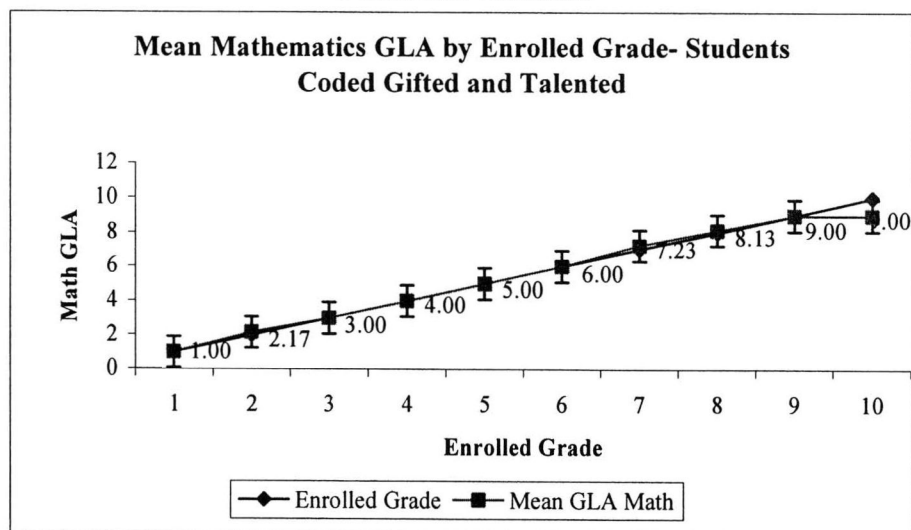
Above line for mean Math GLA is $y = .9319x - .5808$

Gifted and Talented Students



Above mean GLA line is $y = .9281x + 0.3429$

Gifted and Talented



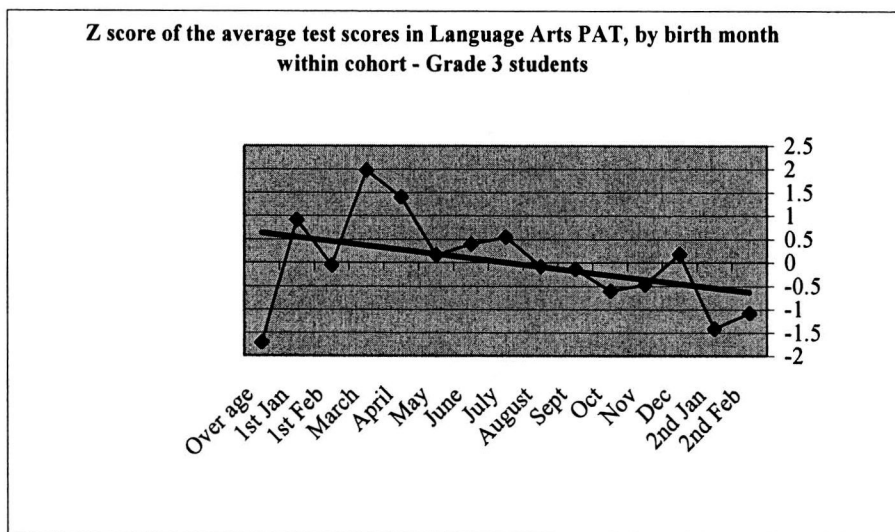
Above line for mean Math GLA is $y = .9467x + .2466$

Discussion

These graphs show that there is a good degree of face validity with the GLA data. For non-coded students the mean GLA in each grade matches the enrolled grade almost perfectly, and this is as expected. One would hypothesize that non-coded students' grade levels of achievement should match very precisely the grade they are enrolled in, and this is what the data show as the mean GLAs range from 0.07 to 0.02 decimals places below the enrolled grades in Math and English. Likewise, one would hypothesize that students with either mild moderate or severe codes mean GLAs would not as precisely reflect the enrolled grade, and again this is what the data show. The mean GLAs in Math and English for students with severe codes range from 1.73 to .28 below enrolled grade, and mild moderate mean GLAs range from 1.56 to .46 below enrolled grade. Finally, for students coded gifted or talented, their mean GLA's in Math and English range from .02 below enrolled grade, to .33 above enrolled grade, which again reflects what one would reasonably expect.

GLA and PAT by Age within Grade Cohorts

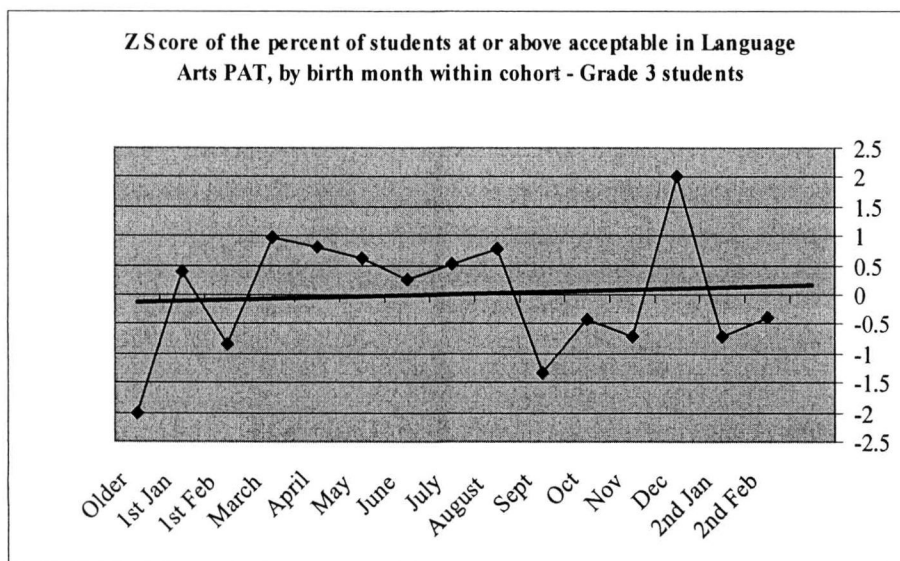
Previous Alberta Learning³ studies have indicated that there is a relative age effect between average PAT scores and birth month within grade cohorts, where older students tend to have higher average test scores than the younger students when measured by z-score average PAT results. The following graph was produced using the z-score Grade 3 English Language Arts PATs for students in the Beyond MIRS dataset.



The above graph shows that the relative age effect noted in the earlier Alberta Learning study is also found within the group of students for whom Beyond MIRS GLA data were collected, that is, older students tend to outperform younger students.

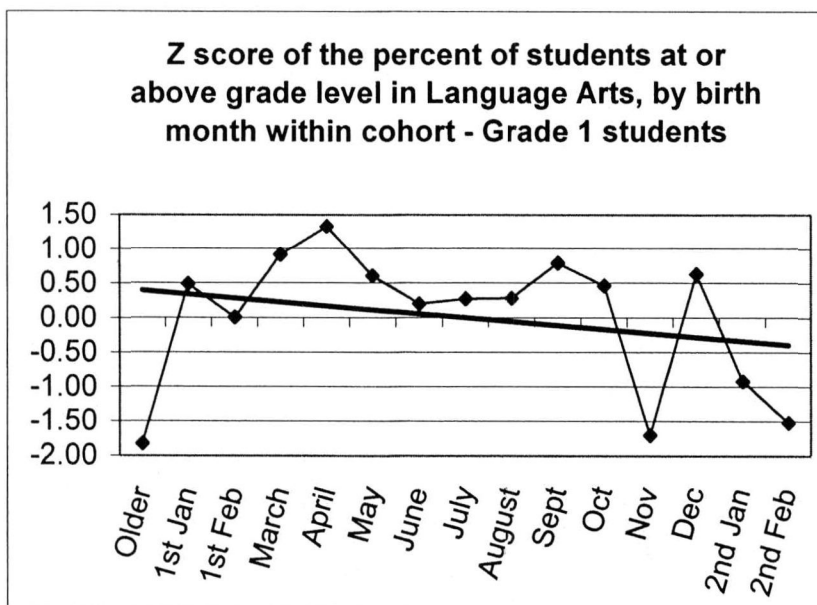
In the following chart, the PAT results for the above students were recoded into “below acceptable” and “at or above acceptable” in order to reflect more direct comparability to the Beyond MIRS GLA data. The percentages of students “at or above acceptable” for each age group were calculated and converted to z-scores. The following graph shows the results of this transformation.

³ *Entry Age, Age Within Cohort, and Achievement*. Alberta Learning, March 2001.



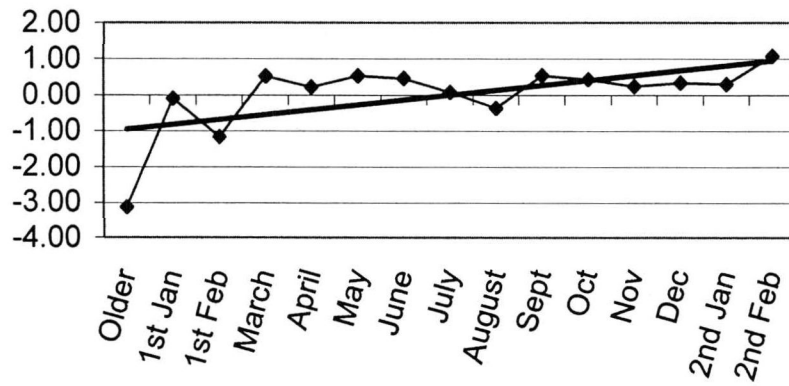
This chart demonstrates that when scores are recoded into “percent at or above acceptable” to mimic the Beyond MIRS GLA data, the relative age effect is no longer apparent.

A comparative analysis was undertaken using GLA data. The percentages of students “at or above” their grade levels in Language Arts and Math were converted to z-scores and plotted. (See graphs below. For complete tables see Appendix 3 in Full Technical Report).

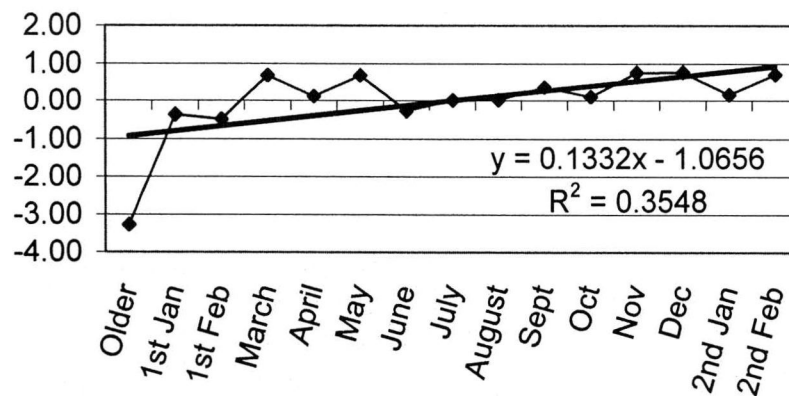


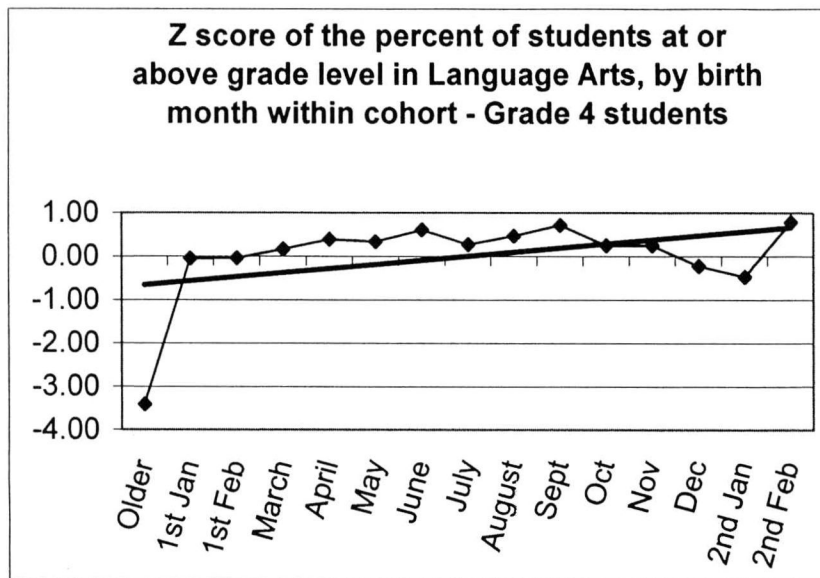
The above graph shows that there is a relative age effect in Grade 1, among Beyond MIRS students, when considering GLA data, similar to the PAT data for grade 3.

Z score of the percent of students at or above grade level in Language Arts, by birth month within cohort - Grade 2 students



Z score of the percent of students at or above grade level in Language Arts, by birth month within cohort - Grade 3 students





The above three charts show that the relative age effect, as measured by GLA data, disappears in Grade 2 and later grades, and even appears to move slightly in the other direction.

Discussion

The relative age effect is apparent when test scores are considered, in this case using the Grade 3 PAT results. Once the scores are coded into two categories, “at or above acceptable” and “below acceptable”, the effect disappears. This is likely due to the fact that older students in the “at or above” group have test scores that are, on average, 2 or 3 points higher than the younger students in the “at or above” group with a subsequent change in the pattern of monthly results. This is important because when we convert PAT data into dichotomous categories, the resultant “line of best fit” mimics the “line of best fit” based on grade 3 GLA dichotomous data.

The Grade 1 GLA, which is based not on scores but on ordinal data quite similar to the “at or above” and “below” construct used in the PATs, does show a relative age effect (which would be more pronounced but for a set of atypical December-born children). This effect disappears in Grade 2 and in following grades. Some of the disappearance likely is due to schools retaining some of the academically weaker younger Grade 1 students, removing them from the Grade 2 cohort, and leaving only the stronger of the younger students to make up the birth month averages. The disappearance may also reflect a truly diminishing relative age effect over time.

GLA by PAT- Comparisons using achievement levels

In order to further examine the relationship between the Beyond MIRS data and PATs, both PAT and GLA data were again re-coded into the dichotomous categories of either "Below Acceptable", or "At or Above Acceptable" for PATs; and "Below Grade Level" or "At or Above Grade Level" for GLA. These were then crosstabulated with the assumption being students who score at or above the acceptable level tend to be at or above grade level, and likewise those that score below acceptable tend to be below grade level, in the majority of cases.

The following tables resulted supporting our hypothesis as 97%-99% of the students who are at grade level are also at or above the acceptable level.

		<i>Grade Level of Achievement – English Language Arts</i>		
		<i>Below Grade Level</i>	<i>At or Above Grade Level</i>	<i>Total</i>
<i>PAT - Grade 3</i>	<i>Below</i>	33.5%	66.5%	546
	<i>Accept.</i>	(183)	(363)	
<i>English Language Arts</i>	<i>Accept. or Excellent</i>	3.0% (133)	97% (4357)	4490
<i>PAT - Grade 6</i>	<i>Below</i>	24.9%	75.1%	650
	<i>Accept.</i>	(162)	(488)	
<i>English Language Arts</i>	<i>Accept. or Excellent</i>	2.3% (107)	97.7% (4482)	4589
<i>PAT - Grade 9</i>	<i>Below</i>	15.6%	84.4%	475
	<i>Accept.</i>	(74)	(401)	

<i>English Language Arts</i>	<i>Accept. or Excellent</i>	.9% (36)	99.1% (4070)	4106
--------------------------------------	-------------------------------------	-------------	-----------------	------

		<i>Grade Level of Achievement – Math</i>		
		<i>Below Grade Level</i>	<i>At or Above Grade Level</i>	<i>Total</i>
<i>PAT - Grade 3</i>	<i>Below</i>	19.3%	80.7%	565
	<i>Accept.</i>	(109)	(456)	
<i>Math</i>	<i>Accept. or Excellent</i>	1.2% (51)	98.8% (4387)	4438
<i>PAT - Grade 6</i>	<i>Below</i>	15.3%	84.6%	600
	<i>Accept.</i>	(92)	(508)	
<i>Math</i>	<i>Accept. or Excellent</i>	.8% (37)	99.2% (4572)	4609
<i>PAT - Grade 9</i>	<i>Below</i>	12.2%	87.8%	915
	<i>Accept.</i>	(112)	(803)	
<i>Math</i>	<i>Accept. or Excellent</i>	.2% (7)	99.8% (3732)	3739

Gamma

All of the above observed relationships were significant when measured by Chi square. Gamma values were subsequently calculated in order to determine the strength of the relationships.

Gamma is a proportional reduction in error (PRE) measure. In short, PREs measure the degree to which knowing the value of the independent variable will reduce error in predicting the value of the dependent variable. GLA was used as the independent measure, with PAT results being set as the dependent. In other words, Gamma provides us with a measure of the degree to which we will be able to predict a student's PAT achievement level, given their GLA.

The formula for Gamma is:

$$\gamma = \frac{N_s - N_d}{N_s + N_d}$$

N_s is the number of similar (concordant) pairs, and N_d is the number of dissimilar (discordant) pairs. To calculate N_s , each cell frequency is multiplied by the sum of the cell frequencies below and to the right of it, and then their products are summed. To calculate N_d , each cell frequency is multiplied by the sum of the cell frequencies above and to their right, and then their products are summed. For example, given the table below for Grade 9 English, Gamma would be calculated as follows:

Gamma- Grade 9 English

		<i>Below Grade Level</i>	<i>At or Above Grade Level</i>	<i>Total</i>
<i>PAT - Grade 3 English Language Arts</i>	Below	15.6%	84.4%	475
	Accept.	(74)	(401)	
	Accept.	.9%	99.1%	4106
	or	(36)	(4070)	
	Excellent			

$$\gamma = \frac{N_s - N_d}{N_s + N_d}$$

$$N_s = 74 \times 4070$$

$$= 301,180$$

$$N_d = 36 \times 401$$

$$= 14,436$$

$$\gamma = \frac{301,180 - 14,436}{301,180 + 14,436}$$

$$\gamma = \frac{286,744}{315,616}$$

$$\gamma = .909$$

In layman's terms, this means knowing a student's Grade 9 English GLA level, gives us roughly a 91% chance of correctly predicting Grade 9 English LA PAT level. However, the above formula has a tendency to overstate the strength of a relationship when any cell has very low values, such as the acceptable PAT but below grade level cell in the grade 9 math data.

The following table lists the Gamma values for the relationships tested⁴.

<i>PAT by GLA- Grade and Subject</i>	<i>Gamma</i>
Gr. 3 Eng., LA	.886
Gr. 6 Eng. LA	.866
Gr. 9 Eng. LA	.909
Gr. 3 Math	.907
Gr. 6 Math	.914
Gr. 9 Math	.973

Analysis of Students Below Grade Level

In the Beyond MIRS pilot, it is possible to compare the ratings given by teachers through the GLA and by a standardized test through the PAT, in Grades 3, 6 and 9. In each case, it is possible to identify the students who are rated as below grade by their teachers (GLA) and those rated as below acceptable standard by the PAT.

One would expect some differences in the designation of individuals in the two ratings, since the teachers have an array of assessments available to do the rating

⁴ A similar analysis was conducted using just jurisdictions other than the main supplier of data. Owing to the smaller n's, it was only possible to calculate Gamma values for Grade 3 and Grade 6 English LA GLA by PAT. The resulting values were .950 and .687 respectively.

whereas the PAT is a single pencil and paper test. However, since the objective of both methods is to measure how well a student is performing as compared to the learning outcomes in the Program of Studies, one would expect an overall positive relationship between the number of students identified as “below” by both methods.

An examination of the Beyond MIRS pilot data shows that this assumption departs most dramatically for grade 9 math within a general pattern where for both Language Arts and Math fewer students are identified as “below” in the GLA ratings than are so identified in the PAT ratings. The following tables illustrate the differences:

Grade 3 ELA	Count	%
Wrote	5036	
Below on PAT	546	10.8%
Below on GLA	316	6.3%

Grade 3 Math	Count	%
Wrote	5003	
Below on PAT	565	11.3%
Below on GLA	160	3.2%

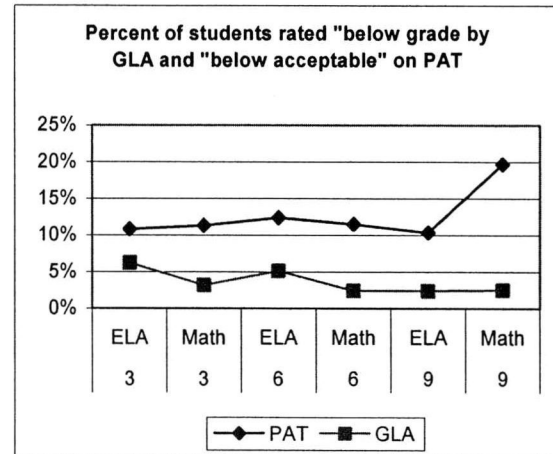
Grade 6 ELA	Count	%
Wrote	5239	
Below on PAT	650	12.4%
Below on GLA	269	5.1%

Grade 6 Math	Count	%
Wrote	5209	
Below on PAT	600	11.5%
Below on GLA	129	2.5%

Grade 9 ELA	Count	%
Wrote	4581	
Below on PAT	475	10.4%
Below on GLA	110	2.4%

Grade 9 Math	Count	%
Wrote	4652	
Below on PAT	915	19.7%
Below on GLA	119	2.6%

The above tables show a difference between the GLA and PAT ratings of “below”, with the gap between the two ratings growing as grade levels increase. The increasing gap can also be shown graphically:



The above analysis seems contrary to the strong Gamma values, and as such, further study was undertaken.

Kendall's tau-b values were calculated in the place of Gamma as a more conservative measure, using the formula:

$$\frac{Ns - Nd}{\sqrt{(Ns + Nd + Tx)(Ns + Nd + Ty)}}$$

where Ns and Nd are the same as Gamma, and Tx designates ties on the independent variable, and Ty designates ties on the dependent variable.

Again using the Grade 9 English values,
 $Ns = 74 \times 4070$
 $= 301,180$

$$\begin{aligned} N_d &= 36 \times 401 \\ &= 14,436 \end{aligned}$$

$$T_x = (74 \times 36) + (401 \times 4070)$$

$$T_x = 1,634,734$$

$$\begin{aligned} T_y &= (74 \times 401) + (36 \times 4070) \\ T_y &= 176,194 \end{aligned}$$

$$\text{Tau-b} = \frac{301,180 - 14,436}{\sqrt{(301,180 + 14,436 + 1,634,734)(301,180 + 14,436 + 176,194)}}$$

$$\frac{286,744}{\sqrt{(1,950,350)(491,810)}}$$

$$\frac{286,744}{\sqrt{959,201,633,500}}$$

$$\frac{286,744}{979,388.4}$$

$$.293$$

All relationships tested were at the $p < .01$ levels meaning they were significant. However, the p-value only shows that the relationships observed did not occur by chance. The tau-b is used as an inferential statistic to show the strength of those relationships. The following table shows all tau-b values for the relationships tested and from this one can conclude that the relationships are moderate in strength.

<i>PAT by GLA- Grade and Subject</i>	<i>Tau-b</i>
Gr. 3 Eng., LA	.392
Gr. 6 Eng. LA	.337
Gr. 9 Eng. LA	.293
Gr. 3 Math	.326
Gr. 6 Math	.298
Gr. 9 Math	.303

Discussion

A primary reason for provincial aggregation of Grade Level of

Achievement data is evaluation of education programs such as special education, English as a Second Language, etc. The GLA by PAT analysis demonstrates that GLA data can indeed supplement PAT data with reasonable reliability and validity for the purposes of program evaluation. This observation is particularly relevant for those grades that do not have PAT testing where GLA can serve as a proxy for PAT data. Additionally, it is useful to be able to supplement PAT data with GLA data in grades 3, 6 and 9 as the added advantage would be broader and richer data to inform program evaluation related decisions.

Further, the fact that the tau-b values show moderate strength lends credibility to the process of collecting GLA. A perfect correlation of 1.0 between GLA and PAT is not an expected nor a desirable condition given the inherent

differences underlying the evaluation designs. PAT data are derived from a single paper and pencil test whereas GLA data are based on numerous and more dynamic observations over time, and thus should be a much richer method of assessment, which one could reasonably assume to produce, positively correlated albeit slightly different data than a PAT result.

GLA and Gender

The 2003 analysis of PISA results⁵ found that females did much better than males in reading, but males tended to outperform females in mathematics. This pattern of gender differentiation is consistent with the general literature on gender-based test performance differences (Pope, Wentzel and Cammaert, 2002). As another test of the concurrent validity of the GLA data, a gender analysis using Mathematics and English Language Arts means was conducted.

Both Mathematics and English Language Arts data were grouped by male and female, according to grade. Each grade's GLA was totaled, and a mean was calculated. The mean differences between males and females were compared using a T-test for means calculation, and the following tables were produced.

English LA GLA T-Tests

Enrolled Grade	Gender	N	Mean GLA	Sig.
1	M	2289	1.00	.394
	F	2203	1.00	
2	M	2759	1.91	.000
	F	2469	1.94	
3	M	2778	2.85	.002
	F	2720	2.88	
4	M	2901	3.80	.001
	F	2693	3.84	
5	M	2869	4.74	.000
	F	2800	4.86	
6	M	2910	5.76	.000
	F	2876	5.85	
7	M	3057	6.80	.000
	F	3076	6.89	
8	M	3094	7.78	.000
	F	2937	7.91	
9	M	2749	8.78	.000
	F	2422	8.88	

Math GLA T-Tests

Enrolled Grade	Gender	N	Mean GLA	Sig.
1	M	2519	1.00	.169
	F	2384	1.01	
2	M	2753	1.96	.418
	F	2468	1.96	
3	M	2739	2.91	.774
	F	2690	2.91	
4	M	2885	3.89	.191
	F	2663	3.90	
5	M	2852	4.84	.000
	F	2768	4.90	
6	M	2876	5.82	.017
	F	2852	5.86	
7	M	3089	6.85	.000
	F	3047	6.91	
8	M	3062	7.85	.000
	F	2914	7.93	
9	M	2734	8.81	.001
	F	2398	8.88	

⁵ PISA 2003 — The 2003 Canadian Report
Measuring Up: Canadian Results of the OECD PISA Study
The performance of Canada's Youth in Mathematics, Reading, Science and Problem Solving
 2003 First Findings for Canadians Aged 15

The above tables show that females outperformed males in English Language Arts by small margins, but the differences were nonetheless statistically significant. The difference between males' and females' mean scores in math were not as pronounced, however they were significant in grades 5 to 9, where females again performed slightly better than males.

Discussion

The results of the gender analysis of GLA data demonstrate concurrent validity with the 2003 PISA gender based results in language arts. However, the GLA math data, while demonstrating no significant differences between males and females in grades 1-4 do demonstrate that females have significantly higher GLA than do males in grades 5-9.

The GLA results for grades 8 and 9 would be most closely comparable to the PISA data for 15 year olds. The GLA and PISA gender analysis in Mathematics are in opposite directions. This appears to suggest the GLA data lack concurrent validity with the PISA data; however, Pope, Wentzel and Cammaert (2002: 284)) studied the relationship between diploma exam scores and the school awarded mark in all diploma exam subjects and found, "For the school-awarded score results, every course that showed statistically significant gender relationships...had results in the direction of girls outperforming boys." The GLA data reported here demonstrate consistent patterns with the school awarded score data reported in the Pope, Wentzel and Cammaert (2002) study, and may support the hypothesis that there may be

"...some sort of differential favoritism in favor of girls in terms of school-awarded scores." These gender relationships are definitely an area worthy of further study both in relationship to GLA data but also in relationship to provincial achievement test data.

Data Availability for Sub-Groups of Coded Students

In order to determine whether there was a relative value to GLA for student reporting for sub-groups of coded students, the PAT achievement component standard levels were recoded into the following categories:

- “At Acceptable”
- “Below Acceptable”
- “Excellent”
- “Missing/Did Not Write”

PAT standard levels were then crosstabulated with the re-coded expanded code variable, resulting in the following, for Grade 3 English Language Arts:

		<i>Student Code Groups</i>				
Grade 3 English Language Arts Achievement Component Standards		No Expanded Codes	Severe (Codes 40 through 49)	Mild/Moderate (Codes 50 through 59)	Gifted and Talented	Total
	At Acceptable	3392	66	152	4	3614
	Below Acceptable	498	16	34	0	548
	Excellent	872	3	1	2	878
	Missing/ Did Not Write	278	62	179	0	519
		5.5% of column total	42% of column total	49% of column total		9% of column total
Total		5040	147	366	6	5559

This indicates that there are no data available for nearly 50% of the severe and mild/moderate sub-populations. English Language Arts GLA data were substituted in the place of PAT data selecting only those students for whom data was previously missing. The following table resulted.

		<i>Student Code Groups</i>			
Grade 3 English Language Arts GLA Levels		No Expanded Codes	Severe (Codes 40 through 49)	Mild/Moderate (Codes 50 through 59)	Total
	Below Grade Level	81	52	166	299
	At Grade Level or Above	187	9	12	208
	Total	268	61	178	507

Substituting Grade 3 English Language Arts GLA data provided data for 98% of the population for whom no PAT data were previously available. Further, when GLA data are substituted, data are unavailable for only .7% of the sub-population coded severe (compared to 42% missing).

Similarly, using GLA means data are unavailable for only .3% of the students in the dataset coded mild/moderate (compared to 49% using PAT data). Similar findings results for other grades and subjects. (See Appendix 4 in Full Technical Report).

Conclusions

This analysis of Beyond MIRS GLA data was undertaken to assess the validity, reliability and ultimately the utility of the GLA data for judging program impacts. The analysis has demonstrated that:

- GLA data, as expected, have a leptokurtic⁶ distribution when applied to the general student population, indicating that most students are achieving at grade level. This is also evident in the Spearman correlations between GLA and enrolled grade.
- GLA for sub-groups such as coded students has a greater distribution and wider variance that increases the utility of the data for judging program impact for these sub-groups (see graphs on pps. 10-12).
- Relationships between GLA and students' age expressed in years-months indicate GLA data have reasonable concurrent validity with PAT data converted to match the format of the GLA data. Variance in GLA data was lower than the variance in PAT data which was also expected given the multiple measures over time that underlie the GLA data.
- The GLA by PAT analysis demonstrates that GLA data can supplement PAT data with reasonable reliability and validity, and with added depth for the purposes of program evaluation. This observation is particularly relevant for those grades that do not have PAT testing where GLA can serve as a proxy for PAT data.
- GLA data provide important data for the approximately 10% of students in grades 3, 6 and 9 who do not write the PATs, thus filling a strategically critical gap in the student achievement database.
- Gender differential analysis of GLA data show a consistent pattern in relationship to 2002 PISA results for reading, but an inconsistent pattern in mathematics. However, the fact that GLA data demonstrate generally higher scores for girls than boys is consistent with a 2002 study that observed consistently higher school awarded marks for girls. GLA data will be an important data source for further study of gender-based achievement.
- Most of the data submitted in the first year of the Beyond MIRS Pilot Project were attributed to a jurisdiction that had acquired considerable experience with GLA reporting. This study and the related conclusions will need to be verified when additional jurisdictions' data are available for analysis.

⁶ As described at http://www.isixsigma.com/dictionary/Leptokurtic_Distribution-268.htm, "A leptokurtic distribution is symmetrical in shape, similar to a normal distribution, but the center peak is much higher; that is, there is a higher frequency of values near the mean [with resultant reduced variation]."

Bibliography

Alberta Learning (2001). *Entry Age, Age Within Cohort, and Achievement*.

Archer, K., Gibbins, R., and Youngman, L. (1998). Explorations: A Navigator's Guide to Quantitative Research in Canadian Political Science

CMEC PISA (2003). *The 2003 Canadian Report Measuring Up: Canadian Results of the OECD PISA Study The performance of Canada's Youth in Mathematics, Reading, Science and Problem Solving 2003- First Findings for Canadians Aged 15*

Pope, G.A., **Wentzel, C.**, & Cammaert, R. (2003). Relationships between gender and Alberta diploma scores. Alberta Journal of Educational Research, 48(4), 275-286